Articles

Toward Clinical Applications of Health Status Measures: Sensitivity of Scales to Clinically Important Changes

Richard A. Deyo and Thomas S. Inui

While the validity and reliability of many newer health status instruments have been reported, few data are available regarding the sensitivity of these instruments to clinically discernible changes in patient status. We studied this feature of the Sickness Impact Profile (SIP) in a group of patients with rheumatoid arthritis, comparing it with more traditional measures of functional status (the American Rheumatism Association (ARA) functional classification and a patient self-rating scale). Four different approaches were devised to measure "sensitivity to clinical change." These involved comparisons of functional score changes with clinical changes in patient status which were independently agreed upon by both clinician and patient, and also comparisons with several clinical disease severity indicators. When applied to groups of patients, the SIP and the patient self-rating scale were modestly superior to the ARA scale, but neither the SIP nor the self-rating scale was clearly superior to the other. For considering individual patients, all of the scales were relatively insensitive, and predictive accuracy for clinically estimated change was low. New strategies for assessing sensitivity to small changes should be developed and applied to health status and functional scales. Attention to this

---

Supported in part by the Health Services Research and Development Program, Seattle Veterans Administration Medical Center and The Robert Wood Johnson Clinical Scholars Program. The views and opinions expressed in this article are those of the authors and may not reflect those of The Robert Wood Johnson Foundation or the Veterans Administration.

At the time of this research, Dr. Deyo was a Veterans Administration Fellow in The Robert Wood Johnson Clinical Scholars Program. Dr. Inui is affiliated with The Robert Wood Johnson Clinical Scholars Program, the Department of Medicine, University of Washington, and Seattle Veterans Administration Medical Center, Seattle, Washington.

Address correspondence and requests for reprints to: Dr. Richard A. Deyo, Department of Medicine, Division of General Internal Medicine, University of Texas Health Science Center, 7703 Floyd Curl Drive, San Antonio, TX 78284.

characteristic should allow refinement of existing scales and may enhance their clinical usefulness.

A new generation of questionnaire instruments for quantifying health and functional status has appeared over the past decade through the combined efforts of clinicians and social scientists [1-5]. For many medical conditions, relevant patient outcomes may include physical, psychological, and social function, requiring the type of global assessment provided by these newer instruments. These health status measures have been successfully applied in large community-based studies [3,6], but it is clear that some investigators intend that they be used in clinical settings as well [4,5,7].

Health status measures might be used in several ways in a clinical setting. They could be used as outcome measures for clinical research or program evaluation, or on a one-time basis, to provide a functional and psychosocial profile of individual patients; or they might be used serially to monitor the natural history of disease or responses to standard interventions. If these instruments are to be used for serial assessments of individual patients, it is important to know how much score variability may occur when the patient is actually clinically stable, and whether the instrument will reflect changes that would be considered clinically important. While the validity and reliability of these instruments have been assessed in clinical settings [4,5,7], few data are available regarding their sensitivity to clinically discernible changes.

Unfortunately, standard methods for assessing this latter performance characteristic are not available. We describe here several ways of examining this question and apply them to one of the newer health status measures, the Sickness Impact Profile (SIP). We have also compared the SIP with two other functional measures: clinician ratings on a 4-point functional scale (the American Rheumatism Association Functional Classification [8]) and a 7-point patient self-rating scale.

We applied these functional scales to a group of patients with rheumatoid arthritis. This condition was chosen for study because it is in many ways prototypical of chronic diseases, it is a relatively common condition, and some short-term lability of disease "activity" occurs, resulting in better or worse patient function.

In rheumatoid arthritis, as in most other clinical illness, there is no "gold standard" for determining what constitutes a clinically important change. Nonetheless, most observers would probably be unwilling to ignore a change for better or worse which patient and physician inde-

pendently agree has occurred. We use such judgments, therefore, as a standard for comparing the functional scales, and call this alteration in status "clinically estimated change."

METHODS

PATIENTS

With informed consent, patients were recruited from two rheumatology specialty clinics, one at the Seattle Veterans Administration medical center and one at the Seattle U.S. Public Health Service hospital. Consecutive patients with routinely scheduled clinic appointments during a 6-month period were invited to participate if they met the following criteria: an established diagnosis of classical or definite rheumatoid arthritis [9], age between 20 and 79 years, ability to read and write English, and no obvious dementia. A total of 79 patients (73 percent of the eligible pool) completed at least one SIP. Most subjects were white (89 percent); and men (56) outnumbered women (23). Mean age was 59 years, and mean disease duration was 12 years.

For 94 percent of these patients, a plurality of all clinic visits during the 6-month study interval were to the arthritis clinic (even after excluding visits solely for gold injections). For two-thirds, the arthritis, orthopedic, and eye clinics were the only clinics attended. It appears, therefore, that rheumatoid arthritis accounted for most of these patients' medical care utilization and that their comorbidity was probably typical of that in a clinical population with similar demographic characteristics.

MEASURES

The Sickness Impact Profile (SIP) is a standardized questionnaire consisting of 136 items grouped into 12 categories. Three of these categories (Ambulation, Mobility, and Body Care and Movement) are aggregated into a "physical dimension." Four other categories are aggregated into a "psychosocial dimension" (Social Interaction, Communication, Emotional Behavior, Alertness Behavior). The five remaining categories are not aggregated in any way (Eating, Work, Sleep and Rest, Household Management, and Recreation and Pastimes). Each item consists of a statement describing a specific dysfunctional behavior, and respondents indicate whether or not each item describes a dysfunction they experience "today" due to their illness. Scores are calculated for the overall instrument, each category, and the two dimensions.

These calculations use predetermined weights based on estimates of the relative severity of each dysfunction [10]. Worse function is reflected by a higher score, better function by a lower score. The weighted scores for each item are summed for all items checked in a given category to yield a raw score, and this is divided by the total possible score for the category to provide a "percentage" score. In this article, all SIP scores are expressed as percentage score. It can be estimated that a 3-point change in percentage score corresponds to a difference of about four or five responses to items of average weight in the questionnaire. The Appendix illustrates scoring of a hypothetical patient's responses.

The American Rheumatism Association (ARA) functional scale is a 4-point ordinal classification with the following definitions of each class:

I. Complete ability to carry on all usual duties without handicaps

II. Adequate for normal activities despite handicap of discomfort or limited motion

III. Limited only to little or none of duties of usual occupation or self-care

IV. Incapacitated, largely or wholly, bedridden or confined to wheelchair, little or no self-care [8].

Patient self-ratings were provided on a 7-point scale of "overall functioning," which asked patients to consider all the areas of function included by the SIP. This was identical to a scale used during validation studies of the SIP [11].

Patient and clinician estimates of changes in a patient's status were made on a 5-point scale. Patients were asked, "Since your last visit to the doctor, do you think you are: much better, slightly better, the same, slightly worse, or much worse?" Clinicians independently answered the same question (with appropriate changes in wording) without knowledge of patient ratings on either the "function" or change scales. Such a scale of change is commonly used in clinical trials of therapy for arthritic conditions [12]. For our analysis, the categories were collapsed to three, and patient-clinician agreements were sought simply for the direction of change. Thus, if patient and clinician both indicated any degree of improvement, the score was 1; if both indicated no change, the score was 2; and if both indicated any degree of worsening, the score was 3. Agreements on this collapsed scale were denoted "clinically estimated change."

RESEARCH ACTIVITIES

At the time of study enrollment, patients were given a copy of the SIP and instructed in its use by a research assistant. Patients were asked to complete the SIP at home as soon as possible and to mail it to the investigator. Additionally, patients were asked to complete an SIP before each arthritis clinic visit for the subsequent 6-month period, but no more frequently than once per month. Subsequent SIPs were delivered to patients by mail, and no further verbal instructions were given. At the end of each SIP were additional questions asking the patient for self-ratings on the 7-point scale of function and the 5-point scale of overall change. A completed SIP was received from the 79 patients for each of 343 patient visits. This represented a response rate of 89 percent for all instances in which SIPs were distributed.

Participating clinicians were the attending physicians, rheumatology fellows, residents, and nurse practitioners who provided routine patient care and monitored gold therapy in the participating clinics. After each clinic visit for which an SIP had been completed by the patient, the provider was asked to make two ratings: the patient's ARA functional class [8], and estimated change from the previous visit on the 5-point scale.

In this nonexperimental study, laboratory data and x-rays were obtained only when they were felt to be necessary for routine care. For study purposes (item 4 in the "approaches" section), laboratory results were used only if they were obtained within 10 days of the date on which a patient completed an SIP. Anatomic stage [8] was estimated for each patient using the most recent available x-rays.

Within the 343 patient visits for which SIP data were obtained, 140 pairs of consecutive visits were recorded, for which complete SIP data, clinician change ratings, and patient change ratings were obtained. In 75 of these visits, patient and clinician agreed on the direction of change in patient status from the previous visit—or on its lack of change. In 56 visits, one indicated some change while the other indicated no change, and in nine cases, the clinician and patient indicated change in opposite directions.

APPROACHES TO MEASURING SENSITIVITY TO CHANGE

We assessed the sensitivity of the functional scales to clinically important change in the four ways summarized here:

First, the mean, standard deviation, and range of score changes

on each scale were computed for the 75 pairs of consecutive visits at which patient and physician agreed that improvement, worsening, or no change ("clinically estimated change") had occurred. This analysis allowed us to determine whether score changes were indeed occurring, whether these changes were in the expected direction (congruent with clinical estimates of change), and what the magnitude of the average score change might be.

Second, correlations were performed between the change agreed upon by patients and physicians, and the change in score for each functional scale. This analysis allowed us to determine which functional scale showed the strongest overall relationship to clinically estimated change, and whether certain subscales of the SIP showed stronger relationships than others.

Third, the sensitivity and specificity of various degrees of functional score change for predicting "clinically estimated change" was calculated. Beginning with standard definitions of sensitivity, specificity, and predictive value, we developed adaptations of these terms (usually applied to 2 × 2 tables) for use with a 3 × 3 table, as shown in Table 1. These calculations are true to the literal definitions of these terms [13]. The characteristics might best be defined as providing answers to the following questions, using score improvement as an example:

- How likely is the functional scale to detect an improvement which has clinically occurred? (sensitivity of a score improvement)
- How likely is the scale to demonstrate no improvement, when the patient is clinically unchanged or worse? (specificity of score improvement)
- Given a score improvement, how likely is this to be clinically correct? (predictive value of a score improvement).

These three characteristics have been calculated for score improvement, score worsening, and lack of score change—with one exception. The specificity of "no score change" would represent the number of score changes observed for patients who are clinically either better or worse, but it would not indicate whether the direction of score changes was the same as the direction of clinically estimated change. The specificity of "no score change" would therefore be a meaningless summary statistic.

This analysis examines the performance of each functional scale as a clinical predictive "test" of individual patient improvement or deterio-

Table 1: Method of Computing Sensitivity, Specificity, and Predictive Value of Functional Scales* for Predicting Clinically Estimated Change

Sickness Impact Profile Score Change	*Clinically Estimated Change*			
	Better	No Change	Worse	
>3.0 point improvement	5	3	2	10
change ≤3.0 points	12	32	8	52
>3.0 point worsening	3	6	4	13
	20	41	14	75

Sensitivity of a score improvement = 5/20 = .25

Sensitivity of a score worsening = 4/14 = .29

Sensitivity of no score change (≤3.0 points) = .78

$$\text{Specificity of a score improvement} = \frac{32 + 8 + 6 + 4}{41 + 14} = .91$$

$$\text{Specificity of a score worsening} = \frac{5 + 3 + 12 + 32}{20 + 41} = .85$$

Predictive value of score improvement = 5/10 = .50

Predictive value of score worsening = 4/13 = .31

Predictive value of no score change = 32/52 = .61

*A change in Sickness Impact Profile (SIP) score of greater than 3 points is used for illustrative purposes. Tabled entries are number of patient visits.

ration. For this purpose, we examined the sensitivity and specificity of 1-, 2-, 3-, or 4-point SIP score changes in predicting clinically estimated change.

Fourth, since the ARA functional scale is an ordinal scale with only 4 points, it was possible to assess the correlation of SIP scores and patient self-ratings with other clinical indicators of disease severity for patients *within* a single ARA category. Significant correlations found within a single ARA category would suggest sensitivity of the SIP or self-rating scale to differences in clinical status not reflected by an ARA rating. For example, at the time of enrollment, 64 patients were categorized as class II on the ARA scale. Considering the initial SIP scores of these patients, only, there was a correlation with disease duration of .20 (p = .054), suggesting that the SIP scores of these patients demonstrated some meaningful variability. The uniform rating of these patients as class II on the ARA scale, however, would not reflect this variability. Thus, regardless of whether or not the SIP was sensitive to

"clinically estimated change" overall, we could assess whether the SIP showed variability among patients at a single point in time which was not reflected by the ARA scale—and whether this variability was biologically meaningful. For this purpose, correlations were calculated between the SIP and patient hematocrit, erythrocyte sedimentation rate (ESR), grip strength, morning stiffness, and anatomic stage, as well as disease duration. Since hematocrit, ESR, and grip are all influenced by age and sex, these demographic characteristics were controlled as covariates in a multiple linear regression. Since the overwhelming majority of patients were rated in ARA Class II (81 percent of patients at the time of enrollment), this was the only category containing enough patients for this evaluation.

RESULTS

Table 2 shows the mean changes in score corresponding to clinically estimated changes. As expected, all of the scales showed a decrease in mean score when improvement was judged to have occurred. In most cases, the smallest mean change was observed when the patient was judged to be unchanged. In the ARA ratings and the psychosocial-dimension score of the SIP, however, the score change observed when the patient was judged unchanged clinically actually exceeded that seen when the patient was judged worse. Of the various SIP scales, the physical dimension shows the greatest score differential among groups. While SIP score changes were small, it appears that for large groups, a score change or difference of even one percentage point on the SIP scale corresponds to a clinically discernible difference. The relatively wide variation in score changes in each category, however, would not allow such a conclusion for an individual patient.

In Table 3, actual correlations are presented between clinically estimated change and change on the functional scales. Only changes in the SIP physical dimension and patient self-rating scale show significant correlations with clinically estimated changes. In this patient group, patients and physicians may be emphasizing physical factors in making clinical judgments of change. Of the functional scales examined, however, the patient self-ratings showed the strongest correlations with clinically estimated change, suggesting greater sensitivity to change than even the SIP or its subscales. This is not surprising, since judgments by patients comprise both the self-rating scale and part of the scale of clinically estimated change.

The sensitivity, specificity, and predictive values of each func-

Table 2: Functional Scale Score Changes Corresponding to Clinically Estimated Changes

	Mean Score Change (Standard Deviation) for Visits at Which Clinician and Patient Agreed on Change		
Functional Scales	*Patient Judged Better (n = 20 visits)*	*Patient Judged Unchanged (n = 41 visits)*	*Patient Judged Worse (n = 14 visits)*
Overall SIP	-0.88 (5.5)	+0.18 (3.3)	+0.58 (4.4)
Physical dimension	-1.16 (6.2)	+0.03 (3.5)	+1.09 (2.7)
Psychosocial dimension	-0.65 (5.2)	+0.58 (4.7)	+0.44 (4.6)
ARA functional class	-0.222 (.647)	+0.077 (0.523)	0 (0)
Patient self-rating	-0.450 (.605)	-0.077 (0.870)	+0.357 (0.842)

Table 3: Correlations between Clinically Estimated Change and Functional Scale Score Changes

Functional Scales	n	*Correlation with Clinically Estimated Change**	p
Change in overall SIP score	75	.159	.09
Change in SIP physical dimension	75	.261	.01
Change in SIP psychosocial dimension	76	.095	.21
Change in ARA functional scale	70	.171	.08
Change in patient self-rated function (7-point scale)	73	.332	.003

*Spearman rank correlation coefficients.

tional scale in predicting clinically estimated change are shown in Table 4. While all of the scales are relatively specific in indicating changes (i.e., score changes occur infrequently in the face of a stable clinical course), they are all relatively insensitive (i.e., a functional score change occurs in only a minority of cases for which change is said to have occurred on clinical grounds). Of the scales examined, the patient self-rating scale shows the greatest sensitivity to clinically estimated improvement. The SIP does appear to "predict" clinical worsening correctly more often than do clinician ratings on the ARA scale, and in this regard it performs as well as the patient self-rating scale. Because improvement or worsening were relatively infrequent in this

Table 4: Performance of Functional Scales as "Tests" for Clinically Estimated Change

Performance Characteristic for Predicting Clinically Estimated Change	*Change in Overall SIP Score*				*Any Change in ARA Class**	*Any Change in Self-rated Function*
	>1.0	*>2.0*	*>3.0*	*>4.0*		
Sensitivity of a score improvement	.50	.30	.25	.25	.33	.40
Sensitivity of a score worsening	.43	.36	.29	.14	.00	.29
Sensitivity of no score change	.51	.68	.78	.80	.72	.54
Specificity of a score improvement	.75	.84	.91	.94	.92	.79
Specificity of a score worsening	.74	.79	.85	.88	.84	.90
Predictive value of score improvement	.42	.40	.50	.62	.60	.42
Predictive value of score worsening	.27	.28	.31	.22	.00	.33
Predictive value of no score change	.72	.67	.61	.57	.55	.50

*ARA Class = American Rheumatism Association functional class, as judged by clinician.

patient group (a low prevalence of change), the predictive value of a "positive test" (a change in score) is seen to be quite low, limiting the usefulness of such changes for following individual patients. Using a score change of 3 or more points as a criterion, the SIP is roughly equal or superior to the other scales in all predictive values, but its performance, nonetheless, is disappointing.

Considering only patients categorized as ARA functional class II, SIP scores showed a significant association with three of seven measures of disease severity and a marginally significant association with a fourth (disease duration, Table 5). Furthermore, all seven correlations are in the expected directions, since hematocrit and grip strength typically decline with more severe inflammatory disease, while the other indicators are expected to increase. This suggests that the SIP is able to distinguish meaningful differences (biologic and other) among patients *within* a single ARA class at a single point in time. On these grounds, then, it may be more sensitive to clinical differences than the ARA scale. By comparison, the 7-point patient self-evaluation scale showed only one significant association with the measures of disease severity, suggesting that by this method, at least, the SIP is more sensitive to some clinical differences than is the patient self-rating scale.

Table 5: Functional Scale Correlations with Clinical Indicators of Disease Severity among Patients Rated as American Rheumatism Association Functional Class II*

Clinical Measure	*Functional Scales*					
	Overall SIP			*7-point Patient Self-rating Scale*		
	n	*r*	*p*	*n*	*r*	*p*
Hematocrit	50	-.30	<.05	49	-.32	<.05
ESR	29	.39	<.05	29	.20	N.S.
Grip strength	19	-.24	N.S.§	18	-.21	N.S.
Morning stiffness	44	.14	N.S.	44	.11	N.S.
Anatomic stage†	59	.06	N.S.	59	-.05	N.S.
Disease duration	64	.20	.054	64	.13	N.S.
Employment status‡	63	$t = 3.0$	<.05	63	$t = 1.1$	N.S.

*Correlations are Spearman rank correlations for morning stiffness, anatomic stage, and disease duration; partial correlations (controlling for age and sex) for hematocrit, erythrocyte sedimentation rate (ESR), and grip strength.

†Anatomic stage: American Rheumatism Association Anatomic Classification [8].

‡For employment status, a *t*-test of group mean functional scores is reported, comparing employed versus unemployed men under age 65.

§N.S. = Not significant ($p > 0.1$).

DISCUSSION

The sensitivity of health indexes to clinically discernible changes or distinctions in health status is often discussed [4,7], but standardized methods for assessing this characteristic have not been defined. Sensitivity to small distinctions has been called "precision" [14], but we have avoided the term since it is used by other authors to denote reproducibility of results [4]. In contrast, methods for assessing the reliability and validity of measurement scales are well described, and in many cases the terms and statistics used are standardized [15].

Comparison of scales with regard to their sensitivity to change is, therefore, difficult. In most clinical settings, no "gold standard" exists for what represents a real change in clinical status. Individual laboratory tests, symptoms, or physical findings may not correspond to overall changes in health or function. For the evaluation of health status

measures, Sackett has suggested that individuals be studied at the time of intake to care, and again "when their therapists judge that a significant change in function has occurred" [4]. We have utilized a variation of this method, requiring that patient and physician both judge, independently, that a change has occurred. This method seems likely to detect clinically discernible but relatively subtle changes in patient status, which for most chronic conditions are more common than dramatic changes. Previous studies of health indexes have demonstrated score changes in patients who progress from inpatient status to outpatient follow-up [4], in patients undergoing total hip replacement [7], and in hyperthyroid patients receiving definitive therapy [16]. *Prima facie*, these all represent fairly dramatic changes in clinical status and may not be typical of the changes observed among ambulatory patients being treated for chronic diseases.

By two of the four methods we have used (methods 1 and 2), patient self-ratings on a 7-point scale appear to be as good an indicator of clinically discernible change—or better—than either patient responses on the SIP or physician ratings of ARA functional class. This may not be surprising, since patient self-ratings are also a component of the "clinically estimated change" scale used for comparison. Using the predictive value of score changes, a 3-point SIP score change in most cases is equal or superior to the other scales. Of the statistics listed in Table 4, we have emphasized predictive value because it addresses the problem most often faced by a clinician in evaluating an individual patient: given a test result, how likely is it to be correct? Judged by the clinical correlations in Table 5, the SIP is again superior to the other scales. Thus, the SIP and patient self-rating scale both apparently are superior to clinicians' ARA ratings in reflecting clinically discernible changes; but in comparing the SIP and the self-rating scale, neither is clearly superior to the other. Unfortunately, all of the scales demonstrated rather low sensitivity and predictive value for clinical changes.

Our scale for clinically estimating change used terms such as "better," "worse," or "the same," so that this was truly a scale of change rather than a single-state measurement. Such criteria of change have been called "transition variables" [17]. Since this scale registered apparent changes in clinical status which were not detected by any of the functional scales, it appears that such transition variables might be a useful component of health indexes which are to be used clinically for serial measurement. In the case of the SIP, for example, patients, after their responses to each category, might be asked: "Considering all the statements on this page, do you think you are better, worse, or the same since your last visit?"

Our study approach for this article might be criticized for several reasons. Given the particular demographics and comorbidities of our patient sample, we cannot be certain that the results would generalize to other samples of rheumatoid arthritis patients or to patients with other chronic diseases. Even within a single well-described population, different observers may consider different factors in judging a change of health status. Important psychosocial changes might be missed by this method if a clinician is unaware of them, or if he/she has made judgments which consider only physical changes. Furthermore, these clinician judgments are not truly independent of patient judgments, since the patient's self-report to the clinician is likely to influence that person's assessment. It is possible that some patients may try to "please the doctor" by saying they are improved, although for study purposes, the patient's self-rating was a written response mailed to the investigators and not seen by the clinician. A clinician's judgments regarding change may reflect knowledge of laboratory data or symptoms which are not paralleled by functional changes. If two different clinicians saw a patient on consecutive visits, the second clinician would have to rely on the medical record, with its various limitations, to assess changes in the physical examination or laboratory measures. In our study, the scale of improvement was intentionally ambiguous and did not specify change in function, symptoms, or other variables, lest this imply a restriction of the scope of considerations primarily to physical aspects of the disease. Such a restriction would have been undesirable since the SIP purports to measure health in a much more global sense. Despite these reservations, few would be willing to dismiss as unimportant a change which patient and provider agree has occurred.

The sensitivity of functional or health status instruments to clinically discernible change is an important attribute if these instruments are to be used in clinical settings for the care of individual patients. The development of new strategies for assessing sensitivity to change should be encouraged. Greater attention to this issue may permit refinement of current functional scales and allow more rational selection of health status and functional instruments for clinical purposes. If health status instruments are to be used for ambulatory patients, they should be sensitive to fairly subtle changes. Our data suggest that while the SIP may appropriately reflect changes for large groups of patients, it is relatively insensitive to change for an individual patient. With regard to this particular characteristic, the SIP appears to offer minimal advantage over patient self-ratings on a unidimensional scale, although it may offer substantial advantages with regard to reliability, validity on other grounds, and informational content [18].

APPENDIX

Summary of SIP Structure and Content, with a Scoring Example

Category	*No. of Items in Scale*	*Maximum Possible Score*	*Hypothetical Patient's Raw Score (Sum of Item Weights for Positive Responses)*	*Hypothetical Patient's Percent Score (Raw Score/Maximum Possible Score)*
Ambulation	12	84.2	47.4	56.3
Body care and movement	23	200.3	38.3	19.1
Mobility	10	71.9	25.8	35.9
Physical Dimension	45	356.4	111.5	31.3
Emotional behavior	9	70.5	25.1	35.6
Social Interaction	20	145.0	71.0	49.0
Alertness behavior	10	77.7	36.4	46.8
Communication	9	72.5	30.7	42.3
Psychosocial Dimension	48	365.7	163.2	44.6
Work	9	51.5	15.5	30.1
Sleep and rest	7	49.9	16.2	32.5
Eating	9	70.5	0.0	0.0
Household management	10	66.8	20.4	30.5
Recreational activities	8	42.2	7.7	18.2
Overall SIP	136	1003.0	334.5	33.3

REFERENCES

1. Gilson, B. S., J. S. Gilson, et al. The Sickness Impact Profile: Development of an outcome measure of health care. *American Journal of Public Health* 65:1304, 1975.
2. Patrick, D. L., J. W. Busch, and M. M. Chen. Toward an operational definition of health. *Journal of Health and Social Behavior* 14:6, 1973.
3. Brook, R. H., J. E. Ware, A. Davies-Avery, et al. Conceptualization and measurement of health for adults in the Health Insurance Study. Overview. *Medical Care* 17:Supplement to No. 7, 1979.
4. Sackett, D. L., L. W. Chambers, A. S. MacPherson, et al. The development and application of indices of health: General methods and a summary of results. *American Journal of Public Health* 67:423, 1977.
5. Parkerson, G. R., S. Y. Gehlbach, E. H. Wagner, et al. The Duke-UNC Health Profile: An adult health status instrument for primary care. *Medical Care* 19:806, 1981.
6. Kaplan, R. M., J. W. Bush, and L. L. Berry. Health status: Types of validity and the Index of Well-Being. *Health Services Research* 11:489. 1976.
7. Bergner, M., R. A. Bobbitt, et al. The Sickness Impact Profile: Development and final revision of a health status measure. *Medical Care* 19:787, 1981.
8. Steinbrocker, O., C. H. Traeger, and R. C. Batterman. Therapeutic criteria in rheumatoid arthritis. *Journal of the American Medical Association* 140:659, 1949.
9. Ropes, M. W., E. A. Bennett, S. Coff, et al. 1958 revision of diagnostic criteria for rheumatoid arthritis. *Bulletin on the Rheumatic Diseases* 9:175, 1958.
10. Carter, W. B., R. A. Bobbitt, M. Bergner, and B. S. Gilson. Validation of an interval scaling: The Sickness Impact Profile. *Health Services Research* 11(4):516-28, 1976.
11. Bergner, M., R. A. Bobbitt, W. E. Pollard, et al. The Sickness Impact Profile: Validation of a health status measure. *Medical Care* 14:57-67, 1976.
12. Kotzin, B. L., S. Strober, E. Engleman, et al. Treatment of intractable rheumatoid arthritis with total lymphoid irradiation. *New England Journal of Medicine* 305:969-76, 1981.
13. Galen, R. S., and S. R. Gambino. *Beyond Normality: The Predictive Value and Efficiency of Medical Diagnoses*. New York: Wiley, 1975.
14. Ware, J. E., R. H. Brook, A. R. Davies, et al. Choosing measures of health status for individuals in general populations. *American Journal of Public Health* 71:620, 1981.
15. *Standards for Educational and Psychological Tests*. Washington, DC: American Psychological Association, 1974.
16. Rockey, P. H., and R. J. Griep. Behavioral dysfunction in hyperthyroidism—Improvement with treatment. *Archives of Internal Medicine* 5:247, 1980.
17. Feinstein, A. R. Indexes and Criteria of Therapeutic Response. In *Clinical Judgement*. Huntington, NY: Krieger Publishing Co., 1967, p. 247.
18. Deyo, R. A., T. S. Inui, J. D. Leininger, and S. S. Overman. Measuring functional outcomes in chronic disease: A comparison of traditional scales and a self-administered health status questionnaire in patients with rheumatoid arthritis. *Medical Care* 21(2):180-92, 1983.